Statistical Evaluation of Median Estimators for Lognormally Distributed Variables

T. B. Parkin* and J. A. Robinson

ABSTRACT

The increased interest in the variability of soil properties is responsible for recent observations that soil variables are not normally distributed but are more closely approximated by the two-parameter lognormal frequency distribution. Statistical methods commonly applied in the estimation of the median of lognormally distributed data, however, are biased or inefficient. The purpose of this study was to evaluate four statistical methods for estimating, from sample data, the median of a lognormal population. The four statistical methods were: (i) the geometric mean (GM), (ii) a bias-corrected form of the geometric mean (BCGM), (iii) a uniformly minimum variance unbiased (UMVU) estimator, and (iv) the sample median (SM). In addition, two techniques for computing confidence limits about the median were evaluated. Monte Carlo simulations from four different lognormal populations were used in these evaluations to determine the efficacy of these methods as a function of both population variance and sample size (n = 4–100). Results of this work indicate that the UMVU estimator and the BCGM estimators are unbiased and yield estimates with the lowest mean square error. An example is provided that illustrates the application of these techniques.

Many of the complex environmental questions faced by society require more precise quantification of environmental variables and processes. Once complicating factor in such environmental studies is the high degree of spatial variability often exhibited by natural variables. Automated data collection and analysis instrumentation has enabled investigators to collect large data sets in an attempt to deal with the problems of high variability. These developments have enabled better determination of frequency distributions.

Many environmental variables exhibit skewed frequency distributions that can be approximated by the lognormal distribution. Unlike symmetric distributions in which the mean and median have the same value, with nonsymmetric distributions, such as the lognormal distribution, the mean and median have different values. When such distributions occur, a choice exists concerning the summary statistic of interest. In a study of epiphytic bacterial populations on leaf surfaces, the median was chosen as the relevant summary statistic (Hirano et al., 1982). The geometric mean has also been used as a measure of central tendency for populations of bacteria in the rhizosphere (Loper et al., 1984) and for bacterial populations in aquatic environments (Greenberg et al., 1985). In contrast, it has been suggested that, for quantification of denitrification N loss from soils, the mean is a more appropriate estimator than the median (Parkin, 1991).

For skewed data the choice of the appropriate summary statistic is important, as it influences the outcome of statistical tests and, therefore, data interpretation (Parkin et al., 1987; Parkin, 1991). A detailed discussion of the power of statistical tests in detecting differences in the mean vs. the median as well as a discussion of criteria for selection of the mean vs. the median is presented elsewhere (Parkin, 1991, 1993; Parkin and Robinson, 1992). In cases where the population median is the appropriate summary statistic, it is important to accurately estimate this quantity from the data. The choice of the optimum estimator is not the only consideration: confidence intervals for the estimator must also be computed.

In previous studies we reported on methods for estimating the mean, variance, and coefficient of variation for lognormally distributed variables (Parkin et al., 1988) as well as on methods for computing confidence limits for the lognormal mean (Parkin et al., 1990). This study extends those findings by evaluating several methods of estimating the population median and confidence limits for the median of a log-normally distributed variable. We report on four methods of estimating the population median from sample data and two techniques for computing confidence limits of the median. These methods were evaluated using four lognormal distributions and across a range of sample sizes representative of those commonly observed in studies of environmental variables.

METHODS

Estimators of the Median

Four estimators of the population median were evaluated: (i) the sample GM, (ii) BCGM, (iii) SM, and (iv) UMVU. Descriptions of the implementation of these estimators are presented below.

The Sample Geometric Mean

The sample geometric mean is the nth root of the product of n samples. This is equivalent to the antilogarithm of the average of the lognormally transformed sample values:

$$\mathrm{GM} = \exp(\bar{y}) \quad [1]$$

where $\bar{y}$ is the arithmetic average of the lognormally transformed sample values.

For large samples drawn from lognormal populations the GM is, approximately, an unbiased estimator of the population median; however, positive bias exists at small sample sizes, with the magnitude of the bias inversely proportional to the sample size (Gilbert, 1987). The bias associated with the GM is given by:

$$\text{bias } (\%) = 100 \times [\exp(\sigma^2/2n) - 1] \quad [2]$$

where σ^2 is the variance of the lognormally transformed population.

Abbreviations: GM, geometric mean; BCGM, bias-corrected geometric mean; UMVU, uniformly minimum variance unbiased; SM, sample median; UCL, upper confidence limit; LCL, lower confidence limit; MSE, mean square error; VAR, variance; CV, coefficient of variation.

T.B. Parkin, USDA-ARS National Soil Tilth Lab., 2150 Pammel Dr., Ames, IA 50011; and J.A. Robinson, 7922-190-MR, the Upjohn Co., Kalamazoo, MI 49001. Received 31 Jan. 1991
*Corresponding author.

Published in Soil Sci. Soc. Am. J. 57:317–323 (1993).

Bias-Corrected Geometric Mean

Gilbert (1987) reported that the BCGM can be an estimator of the population median if σ^2 is known. He suggested that the variance of the lognormally transformed sample can be used if σ^2 is not known; however, an evaluation of this technique was not done.

The BCGM is calculated according to:

$$\text{BCGM} = \exp(\bar{y} - \hat{\sigma}^2/2n) \quad [3]$$

where $\hat{\sigma}^2$ is the variance of the lognormally transformed sample data.

The Sample Median

The sample median is the center value of the ordered sample data. If an odd number of samples is available, then the median is the sample corresponding to the $[(n + 1)/2]$th rank. If n is even, then the median is calculated as the average of the two center values of the ranked data:

$$\text{SM} = x_{[(n+1)/2]} \qquad n = \text{odd} \quad [4]$$

$$\text{SM} = (x_{(n/2)} + x_{[(n+2)/2]})/2 \qquad n = \text{even} \quad [5]$$

The SM has been reported to be an asymptotically unbiased estimator of the population median (Kleijnen, 1987).

The Uniformly Minimum Variance Unbiased Estimator

The UMVU estimator of the median was developed by Bradu and Mundlak (1970) based on the work of Finney (1941), and is theoretically an unbiased minimum variance estimator:

$$\text{UMVU} = \exp(\bar{y})\psi_n\{-\hat{\sigma}^2/[2(n-1)]\} \quad [6]$$

where ψ_n is the power function:

$$\psi_n(z) = 1 + \frac{z(n-1)}{n} + \frac{z^2(n-1)^3}{n^2(n+1)2!} + \frac{z^3(n-1)^5}{n^3(n+1)(n+3)3!} + \cdots \quad [7]$$

where $\{-\hat{\sigma}^2/[2(n-1)]\}$ is substituted for z in Eq. [6], and n is the sample size.

In this study, Eq. [7] was evaluated until the final term accounted for $<1\%$ of the sum of the preceding terms. This usually required from six to 10 terms.

Confidence Interval Estimators of the Median

Two methods for constructing a confidence interval for the median of a lognormally distributed variable were evaluated. A description of the methods is presented below.

Method 1

This method is based on the normal approximation of the binomial distribution in which order statistics are calculated to yield an approximate 95% confidence interval (Snedecor and Cochran, 1967, p. 123–125.) Although this technique is approximate, it is simple and straightforward to apply since it does not necessitate the use of the binomial distribution (Conover, 1980, p. 493). Upper and lower confidence limits are given by:

$$\text{LCL}_1 = x(s) \quad [8]$$

$$\text{UCL}_1 = x(r) \quad [9]$$

where $x(s)$ and $x(r)$ are the sth and rth values of the ordered sample data. The values of r and s are calculated according to:

$$s = (n + 1)/2 - \sqrt{n},$$

rounded down to the nearest integer [10]

$$r = (n + 1)/2 + \sqrt{n},$$

rounded up to the nearest integer [11]

Method 2

This technique is the asymptotic or normal theory method applied to lognormally transformed sample data (Gilbert, 1987). Lower and upper confidence limits are given by:

$$\text{LCL}_2 = \exp(\bar{y} - t_{0.975,n-1}\sqrt{\hat{\sigma}^2/n}) \quad [12]$$

$$\text{UCL}_2 = \exp(\bar{y} + t_{0.975,n-1}\sqrt{\hat{\sigma}^2/n}) \quad [13]$$

where t = the critical value from Student's t distribution with $n - 1$ degrees of freedom for a two-tailed value of $\alpha = 0.05$.

Evaluation of Estimators

Evaluation of Median Estimators

Two criteria were used to evaluate the estimators of the median: bias and MSE. A biased estimator is one whose expected value deviates from the population parameter being estimated. Thus, a biased estimator will, on average, either underestimate or overestimate the population parameter. Bias alone is not a sufficient criterion to choose a method, as the variance associated with the estimator may be so large as to limit its utility. When bias exists, however, variance alone is not a valid criterion (Barnett, 1973). In this situation the MSE is used, with the optimal estimator being the one with the lowest MSE. The MSE is the sum of the variance and the bias squared:

$$\text{MSE} = \text{VAR} + \text{bias}^2 \quad [14]$$

Monte Carlo simulation was used to determine the influence of sample size on bias and MSE of the median estimators. Sample sizes of four to 100 observations (incremented by two) were selected from each of four lognormal populations with known properties. The four lognormal distributions used spanned the range of positively skewed distributions observed for soils data (Parkin et al., 1988). Each population had a mean of 10 and coefficients of variation of 50, 100, 200, and 500%. The statistical properties of these two-parameter lognormal populations are given in Table 1. These same distributions were used in our previous study to evaluate methods for estimating the mean, variance, coefficient of variation, and confidence intervals of the mean of lognormally distributed data (Parkin et al., 1988; Parkin et al., 1990).

Twenty-five thousand Monte Carlo simulations were run for each sample size using an algorithm for generating random variates from a lognormal distribution, and the population median was estimated by the four methods for each simulation. The pseudo-random number generator resident in the programming language (Turbo Basic, Borland, Scotts Valley, CA) was used in these simulations. The bias, bias percentage, and var-

Table 1. Statistical properties of the four test populations.

Population	Mean	Median	Mode	Skewness	Variance	Mean of logarithms	Standard deviation of logarithms
A	10.0	8.944	7.155	1.625	25.0	2.191	0.4734
B	10.0	7.071	3.536	4.000	100.0	1.956	0.8326
C	10.0	4.472	0.894	14.000	400.0	1.498	1.2690
D	10.0	1.961	0.075	140.000	2500.0	0.674	1.8050

iance of the 25 000 simulations for each sample size were calculated by:

$$\text{Bias} = \frac{1}{n_s}\sum_{i=1}^{n_s}(p_i - P) \qquad [15]$$

$$\text{Bias (\%)} = 100 \times \text{Bias}/P \qquad [16]$$

$$\text{Variance} = \frac{1}{n_s}\sum_{i=1}^{n_s}(p_i - p)^2 \qquad [17]$$

where

n_s = the number of Monte Carlo simulations = 25 000,
P = the true value of the population median,
p_i = the estimate of the population median for the ith Monte Carlo simulation obtained by one of the four methods, and
p = the average of medians estimated from the n_s Monte Carlo simulations.

The MSE for each estimator was calculated according to Eq. [14].

Confidence Interval Evaluations

The confidence interval methods were evaluated with regard to efficacy in providing coverage at the stated level. Confidence intervals were constructed for each stimulation at an α level of 0.05, with the error divided equally between the two tails. This is equivalent to constructing two one-sided 97.5% confidence limits, thus yielding a two-sided 95% confidence interval. The proportion of times the actual population median was less than the UCL was counted (97.5% expected for an α = 0.05). Likewise, the proportion of times the population median was greater than the lower limit was counted (97.5% expected for an α = 0.05). The results of this analysis are estimates of the actual probability levels for the UCL and LCL calculated by each method. The methods we judged by how well these estimated probability limits compared with the theoretical 97.5% target level. These evaluations were performed for sample sizes of four to 100 at increments of two (25 000 simulations at each sample size) drawn from each of the four lognormal populations.

RESULTS AND DISCUSSION

Estimators of the Median

For all populations and across the range of sample sizes tested, both the GM and the SM estimators yield positively biased estimates of the true population median. That is, they produce values which, on the average, overestimate the true population median (Table 2). The degree of bias increases with increasing population variance, and decreases with increasing sample size. For a sample size of four, bias of the GM and SM is not severe for Population A (2.9 and 5.1%, respectively). As population variance increases, bias also increases, such that

Table 2. The bias percentage of the median estimators at select sample sizes. Results for all four estimators were obtained from the Monte Carlo simulations, and in addition the bias percentage associated with the geometric mean was calculated using Eq. [2] using the known population variances.

	Sample size	Bias percentage: Geometric mean (Eq. [2])	Geometric mean	Sample median	Uniformly minimum variance unbiased	Bias-corrected geometric mean
	no.	%				
Population A	4	2.89	2.81	5.14	−0.07	−0.21
	12	0.93	0.94	1.53	0.01	−0.04
	20	0.62	0.62	0.96	0.0	−0.05
	40	0.28	0.26	0.48	−0.02	0.03
	100	0.11	0.12	0.22	0.01	0.0
Population B	4	9.05	8.67	16.4	−0.72	0.29
	12	2.93	2.82	4.69	−0.13	−0.04
	20	1.75	1.89	2.95	0.10	−0.02
	40	0.87	0.79	1.21	−0.09	−0.02
	100	0.35	0.37	0.58	0.02	−0.02
Population C	4	22.3	22.5	43.4	−1.19	1.62
	12	6.94	6.76	10.9	−0.40	0.54
	20	4.11	3.78	6.39	−0.39	0.12
	40	2.02	1.96	3.18	−0.09	−0.02
	100	0.81	0.71	1.15	−0.10	−0.16
Population D	4	50.3	49.9	108.0	−7.50	4.21
	12	14.5	13.8	24.1	−1.59	0.60
	20	8.49	9.02	14.6	0.14	−0.04
	40	4.16	4.10	6.59	−0.15	0.06
	100	1.64	1.64	2.65	−0.02	0.12

for Population D (CV = 500%) the bias associated with a sample size of four is 50 and 108% for the GM the SM, respectively. Across all the sample sizes evaluated, the bias associated with the SM is approximately twice that of the GM.

Monte Carlo simulation was applied to the UMVU estimator to evaluate the random error associated with the computer simulations. For Population A, the random error associated with the computer simulations was low (±0.07% at n = 4). The random error associated with the simulations was highest for Population D at small sample sizes (7.5% for n = 4). This random error, however, is small compared with the magnitude of the bias percentage exhibited by the SM and GM estimators. An additional check on the simulations is provided by comparing the theoretical bias associated with the GM, calculated from Eq. [2], to results obtained from the Monte Carlo simulations. Across all populations and sample sizes, the theoretical bias percentage and the simulation results for the GM agree well, indicating that random errors associated with the computer simulations were minor.

The bias associated with the BCGM was minimal and fluctuated around zero. Thus, within the error limits of the simulations, the BCGM appears to be an unbiased estimator of the population mean.

Evaluation of the estimators was not based solely on bias; the variance associated with each estimator was also considered. Mean square error is a combined measure of the variance of the estimator as well as the bias (Eq. [8]). For each population, the estimators exhibit a similar trend of decreasing MSE with increasing sample size (Fig. 1). For any given sample size, the SM has the highest MSE. Differences in MSE associated with the estimators are most apparent at small sample sizes (n < 12), and for larger sample sizes (n > 20) only minor differences are observed. For Populations A and B (Fig. 1A and 1B) the MSE associated with the UMVU and GM are nearly identical across the range of sample sizes. For Populations C and D (Fig. 1C, 1D), however, greater differences in MSE associated with the GM and UMVU exist at small sample sizes. Across the range of sample sizes and populations, the MSE associated with the BCGM is nearly identical to the MSE of the UMVU estimator.

Based on these results, we recommend either the UMVU or the BCGM as the estimators of choice for the population median. This is a blanket recommendation made across the range of sample sizes and populations investigated in this study. Since the BCGM and UMVU estimators yield nearly identical results with regard to bias and MSE, the BCGM may be more desirable, as it does not require evaluation of Eq. [7], and is therefore numerically less cumbersome to implement. The GM, while slightly easier to calculate (the bias correction does not have to be made), may show substantial bias at small sample sizes when population variance is high. For large sample sizes (n > 20 Populations A and B; n > 40 Populations C and D) the bias correction may not be necessary; however, this correction is very simple to implement. The SM is never recommended when it is assured that the underlying population is adequately modeled by a lognormal distribution.

Confidence Intervals about the Median

Two methods were applied for calculation of confidence limits about the median. The first technique is a nonparametric method based on the order statistics, which gives an ≈95% confidence interval regardless of the form of the underlying frequency distribution. The second technique is the asymptotic or normal theory method

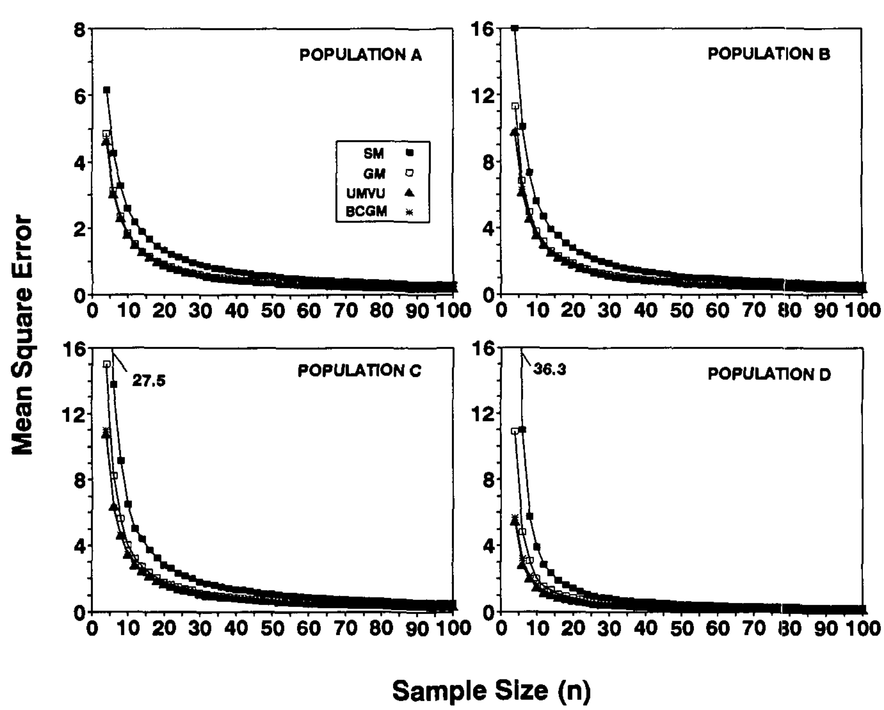

Fig. 1. Mean square error of four estimators of the median as a function of sample size for four lognormal populations. Across all sample sizes and populations, mean square error of bias-corrected geometric mean and uniformly minimum variance unbiased estimator were nearly identical.

applied to lognormally transformed sample data. This technique, when applied to lognormally transformed sample data, yields exact upper and lower 97.5% confidence limits.

Results for the LCL are shown in Fig. 2. The behavior of the two methods as a function of sample size showed similar patterns for all four test populations. The nonparametric method exhibited wide fluctuations, and consistently yielded confidence limits that were greater than the target $(1 - \alpha)/2$ level of 0.975. In contrast, the parametric method showed minor fluctuations centered around the 0.975 level.

Upper confidence limits calculated by the two methods behaved in much the same manner as the LCLs (Fig. 3). The nonparametric method provided coverage >0.975, while the parametric method provided coverage at the target 0.975 level. The overall result of this evaluation is that the nonparametric method yields a 96.8% confidence interval (range 95.4–99.2%), while the parametric method yields a confidence interval at the stated $(1 - \alpha)$ level of 95% (range 94.6–95.4%).

In addition to providing exact UCLs and LCLs at the stated probability level, the parametric method also had a resulting confidence interval that was narrower than the confidence interval provided by the nonparametric method (Table 3). The greatest differences in confidence

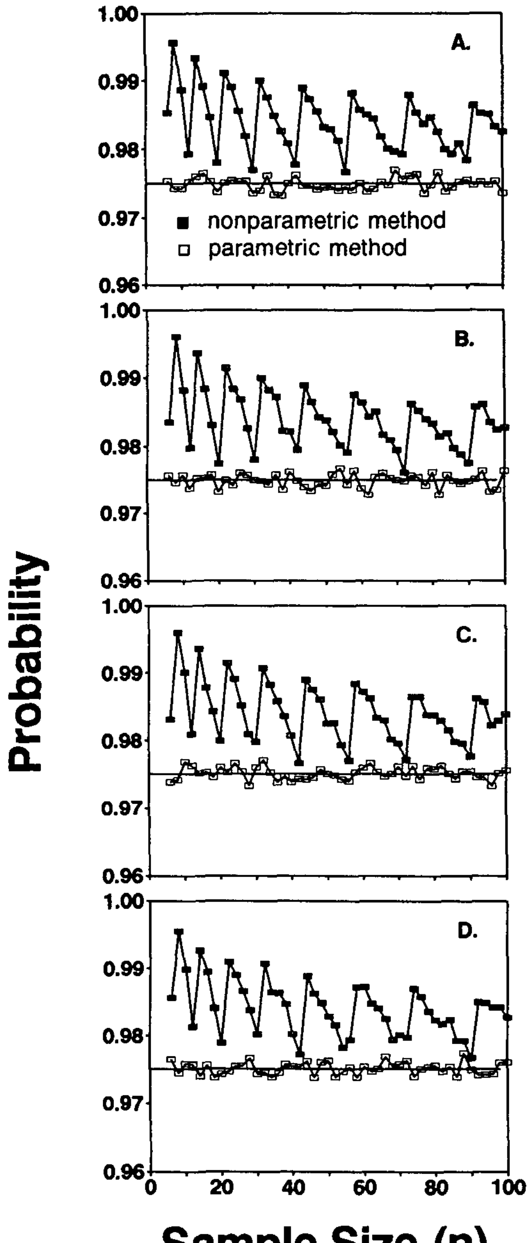

Fig. 2. Realized probability level $(1 - \alpha/2)$ of lower confidence limits calculated by the nonparametric method and the parametric method for each of the four lognormal test populations. The horizontal line in each panel indicates the theoretical 0.975 probability level.

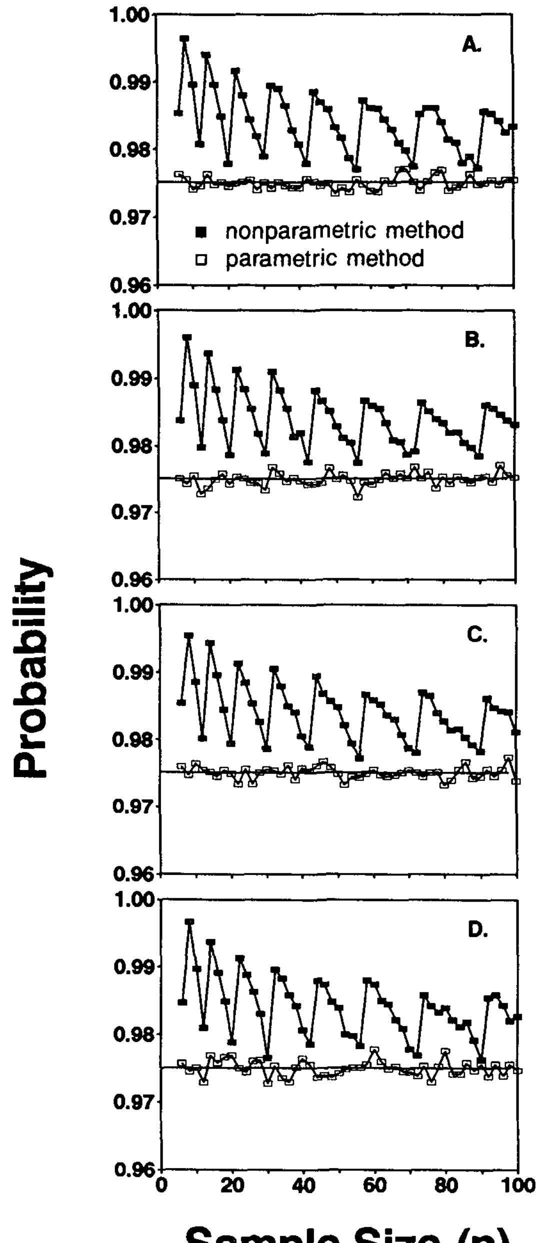

Fig. 3. Realized probability level $(1 - \alpha/2)$ of upper confidence limits calculated by the nonparametric method and the parametric method for each of the four lognormal test populations. The horizontal rule line in each panel indicates the theoretical 0.975 probability level.

interval widths exhibited by the two methods were manifested at low sample sizes ($n < 20$), and for the highly skewed populations (Populations C and D).

Based on these results, we recommend the parametric method for computing confidence limits about the median, as this method provides exact limits and the resulting confidence intervals are narrower than those provided by the nonparametric method. This recommendation is only made when it is assured that the sample data are from a lognormal distribution. It should be noted that, while the parametric method is the optimal method for true lognormal data, the nonparametric method yielded reasonably accurate confidence limits, and thus may be preferred when the underlying population distribution is unknown.

The "saw-tooth" behavior of the results obtained by the nonparametric method is a result of the rounding operations required to compute the confidence limits. We recognize that better methods for computing nonparametric confidence intervals exist. As an alternative to rounding up or down as prescribed by Eq. [10] and [11], exact probabilities could have been computed using the binomial theorem (Conover, 1980, p. 493). Our interest in this study, however, was to evaluate the performance of this simple method in relation to the exact parametric approach.

The purpose of this study was to evaluate median estimators and confidence interval methods applied to lognormal data. Our results indicate that, when the sample data are from a lognormal distribution, the BCGM performs nearly as well as the UMVU estimator, yet is simpler to compute. Also, the parametric method of calculating confidence limits is exact. While the lognormal distribution may be an appropriate model for many soil processes (Parkin et al., 1988), there are situations in which the data, though skewed, may not be lognormally distributed. The robustness of these methods to deviations from lognormality has not been evaluated and in some cases it could be possible that the sample median and the nonparametric confidence limit method would outperform the other methods of this study when the underlying population is not truly lognormal. If the data analyst is not comfortable with the assumption of lognormality, or if the sample size is large enough, the sample median may be used as an estimator, along with the nonparametric method for computation of confidence limits. It must be recognized that the sample median, while asymptotically unbiased, exhibits significant bias at small sample sizes when applied to lognormal data.

Table 3. Average widths of intervals calculated by two confidence interval methods for selected sample sizes.

	Sample size	Nonparametric method	Parametric method
	no.	Interval width	
	6	12.0	9.0
	12	6.96	5.37
Population A	20	5.12	3.96
	40	3.56	2.67
	100	2.26	1.68
	4	20.9	14.5
	12	10.7	7.97
Population B	20	7.49	5.70
	40	5.09	3.83
	100	3.16	2.35
	4	31.4	18.4
	12	12.0	8.46
Population C	20	7.87	5.79
	40	5.11	3.77
	100	3.09	2.29
	4	42.7	18.3
	12	16.8	7.29
Population D	20	5.78	4.01
	40	3.44	2.46
	100	2.01	1.46

APPENDIX

While it is impossible to evaluate the efficacy of estimators using only sample data, the following discussion is provided to illustrate the mechanics of applying the four estimators of the median and two confidence limit methods to sample data. This example data set (Table 4) contains 20 observations of pore water velocity, which were presented by Warrick and Nielsen (1980). This data set has been reported to be lognormally distributed. Table 4 shows the nontransformed and the lognormally transformed pore water velocity data for each observation. Also shown are values of the mean and variance of the lognormally transformed sample data ($\bar{y}$ and $\hat{\sigma}$, respectively), which are used in the following calculations.

Estimators of the Median

The Geometric Mean

Applying Eq. [1] the geometric mean is calculated as:

$$GM = \exp(2.12) = 8.33$$

The Bias-Corrected Geometric Mean

Applying Eq. [3], the BCGM is calculated as:

$$BCGM = \exp(2.12 - 1.90/40) = 7.94$$

The Sample Median

Since there are an even number of samples, Eq. [4] is applied to calculate the SM, which is the average of the 10th and 11th sample values.

$$SM = (7.39 \pm 8.41)/2 = 7.90$$

Table 4. Twenty observations of pore water velocity (v_o) from Table 13.1 of Warrick and Nielsen (1980).

Sample	v_O	$\ln(v_o)$
	cm/d	
1	0.66	−0.4155
2	1.16	0.1484
3	1.52	0.4187
4	2.29	0.8286
5	2.94	1.0784
6	3.46	1.2413
7	4.35	1.4702
8	4.95	1.5994
9	6.11	1.8099
10	7.39	2.0001
11	8.41	2.1294
12	10.38	2.3399
13	12.81	2.5502
14	15.64	2.7498
15	18.64	2.9253
16	24.53	3.1999
17	32.14	3.4701
18	47.47	3.8601
19	88.23	4.4780
20	90.75	4.5081
		$\bar{y} = 2.12$
		$\hat{\sigma}^2 = 1.90$

The Uniformly Minimum Variance Unbiased Estimator

The UMVU estimator of the median is obtained from Eq. [6]. This requires that the ψ_n function (Eq. [7]) be evaluated for $\{-\hat{\sigma}^2/[2(n-1)]\}$. Using data from this example,

$$-\hat{\sigma}^2/[2(n-1)] = -1.90/[2(19)] = -0.03052.$$

Substituting this value for z in Eq. [7] yields a value of 0.9525, and applying this to Eq. [6] yields:

$$\text{UMVU} = \exp(2.12)0.9525 = 7.94$$

With this data set, it is observed that the UMVU and BCGM estimators yield values that are identical. The SM yields an estimate of the median that is slightly lower, while the GM yields a greater estimate for the population median. Thus a choice exists in the selection of a estimate of the underlying population median. The efficacy of an estimator, however, cannot be made based on sample data alone: additional information is required. The results of our study indicate that GM is a biased estimator, and that the best estimators of the population median (i.e., zero bias and lowest MSE) are the UMVU and the BCGM estimators. Thus, for this example, the best estimate of the population median is 7.94.

Confidence Intervals

Parametric Method

Applying Eq. [12] and [13] to the data of Table 4 yields LCLs and UCLs according to the parametric method:

$$\text{LCL} = \exp(2.12 - 2.086\sqrt{1.90/20}) = 4.38$$

$$\text{UCL} = \exp(2.12 + 2.086\sqrt{1.90/20}) = 15.84$$

Nonparametric Method

As described by Eq. [8] and [9], the approximate 97.5% UCL and LCL are the sth and rth values of the ordered data set, where s and r are calculated according to Eq. [10] and [11]. For this example data set, s and r are computed as follows:

$s = (20 + 1)/2 - \sqrt{20} = 6.03$, rounding down to the nearest integer yields $s = 6$.

$r = (20 + 1)/2 + \sqrt{20} = 14.97$, rounding up to nearest integer yields $r = 15$. Thus, the UCL and LCL are 4.35 and 18.64, the 7th and 18th data values of Table 4.

It is observed that the confidence interval width of the parametric method is shorter than that yielded by the nonparametric method. It is impossible to determine from the sample data, however, the efficacy of the confidence limit methods with regard to coverage of the confidence limits provided by the two methods. Based on the simulation results of our study, the parametric method yields exact limits at the 97.5% level.

REFERENCES

Barnett, V. 1973. Comparative statistical inference. John Wiley & Sons, New York.

Bradu, D., and Y. Mundlak. 1970. Estimation in lognormal linear models. J. Am. Stat. Assoc. 65:198–211.

Conover, W.J. 1980. Practical nonparametric statistics. John Wiley & Sons, New York.

Finney, D.J. 1941. On the distribution of a variate whose logarithm is normally distributed. J. Royal Stat. Soc. Suppl. 7:155–161.

Gilbert, R.O. 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold Co., New York.

Greenberg, A.L., R.R. Rhodes, and L.S. Clesceri (ed.). 1985. Standard methods for the examination of water and wastewater. 16th ed. Am. Publ. Health Assoc., Washington, DC.

Hirano, S.S., E.V. Nordheim, D.C. Arny, and C.D. Upper. 1982. lognormal distribution of epiphytic bacterial populations on leaf surfaces. Appl. Environ. Microbiol. 44:695–700.

Kleijnen, J.P.C. 1987. Statistical tools for simulation practioners. Marcel Dekker, New York.

Loper, J.E., T.V. Suslow, and N.M. Schroth. 1984. Lognormal distribution of bacterial populations in the rhizosphere. Phytopathology 74:1454–1460.

Parkin, T.B. 1991. Characterizing the variability of soil denitrification. p. 213–228. *In* N.P. Revsbech and J. Sorensen (ed.) Denitrification in soil and sediment. Plenum Press, New York.

Parkin, T.B. 1993. Evaluation of statistical methods for determining differences between lognormal populations. Agron. J. (in press).

Parkin, T.B., S.T. Chester, and J.A. Robinson. 1990. Calculating confidence intervals for the mean of a lognormally distributed variable. Soil Sci. Soc. Am. J. 54:321–326.

Parkin, T.B., J.J. Meisinger, S.T. Chester, J.L. Starr, and J.A. Robinson. 1988. Evaluation of statistical methods for lognormally distributed variables. Soil Sci. Soc. Am. J. 52:323–329.

Parkin, T.B., and Robinson. 1992. Analysis of lognormal data. Adv. Soil Sci. 20:193–235.

Parkin, T.B., J.L. Starr, and J.M. Meisinger. 1987. Influence of sample size on measurement of soil denitrification. Soil Sci. Soc. Am. J. 51:1492–1501.

Snedecor, G.W. and W.G. Cochran. 1967. Statistical Methods, 6th Edition. The Iowa State Univ. Press, Ames.

Warrick, A.W., and D.R. Nielsen. 1980. Spatial variability of soil physical properties in the field. p. 319–344. *In* D. Hillel (ed.) Applications of soil physics. Academic Press, New York.